DOCUMENT RESUME

ED 162 982 95 SP 013 369

AUTHOR Clark, Christopher M.; And Others

TITLE Identifying Cues for Use in Studies of Teacher Judgment.

INSTITUTION Michigan State Univ., East Lansing. Inst. for Research on Teaching.

SPONS AGENCY National Inst. of Education (DHEW) Washington, D.C. Teaching Div.
PUB DATE Jul 78
CONTRACT 400-76-0073
NOTE 30p.
AVAILABLE FROM The Institute for Research on Teaching, 252 Erickson Hall, Michigan State University, East Lansing, Michigan 48824 ($1.75)

EDRS PRICE MF-$0.83 HC-$2.06 Plus Postage.
DESCRIPTORS *Cognitive Processes; *Decision Making; Educational Research; *Factor Analysis; Multiple Regression Analysis; *Policy Formation; Student Motivation; *Teacher Attitudes

ABSTRACT

Identification of cues on the major factors affecting judgment is discussed in this paper. The advantages and weaknesses of five approaches to identifying, selecting, and defining features of objects or activities to be judged is examined. These features are: (1) logical specification; (2) expert opinion; (3) prespecification and narrowing of a large number of potential cues; (4) allowing cues to emerge during a judgment task, and (5) participant observation in a naturalistic setting. The paper describes a variation of the fourth method for descriptive research on teaching judgment that represents a simple, realistic compromise between the representativeness of cues identified and the time and effort required to identify them. Features relating to students, subject matter, the teacher, and learning environment were identified by teachers as influential. Of the four categories, features relating to student motivation and involvement were mentioned most frequently. (Authors/JD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ERIC
CP
ED162982
U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY
DEC 15 1 38 PM 278
SP013 369

Institute for Research on Teaching

College of Education — Michigan State University

Research Series No. 23

IDENTIFYING CUES FOR USE IN STUDIES OF TEACHER JUDGMENT

Christopher M. Clark,
Robert J. Yinger, and Susan C. Wildfong

Published By

The Institute for Research on Teaching
252 Erickson Hall
Michigan State University
East Lansing, Michigan 48824

July 1978

The work reported herein is sponsored by the Institute for Research on Teaching, College of Education, Michigan State University. The Institute for Research on Teaching is funded primarily by the Teaching Division of the National Institute of Education, United States Department of Health, Education, and Welfare. The opinions expressed in this publication do not necessarily reflect the position, policy, or endorsement of the National Institute of Education. (Contract No. 400-76-0073)

Contents

Current Approaches to Cue Identification..... 2

- Logical Specification..... 3
- Expert Opinion..... 4
- Prespecification and Narrowing of Potential Cues..... 5
- Allowing Cues to Emerge During a Judgment Task..... 7
- Participant-Observation.....10
- Summary.....11

A Study of Teacher Judgment.....13

- Method.....13
- Results.....15
- Feature Configuration.....19
- Discussion.....19
- Limitations.....21

Conclusion.....22

References.....23

Tables

1. Features of Language Arts Activities Identified As Influences on Teacher Judgment.....15
2. Frequency of Reported Use of Activity Features for Each Teacher.....18

Abstract

Identification of cues or the major factors affecting judgment is an important first step in policy-capturing studies of human judgment. In this paper, the authors discuss the advantages and weaknesses of five approaches to identifying, selecting, and defining features of objects or activities to be judged: (1) logical specification, (2) expert opinion, (3) prespecification and narrowing of a large number of potential cues, (4) allowing cues to emerge during a judgment task, and (5) participant observation in a naturalistic setting. The remainder of the paper describes a variation of the method (4) above for descriptive research on teaching judgment that represents a simple, realistic compromise between the representativeness of cues identified and the time and effort required to identify them. This approach is illustrated by a study of teacher judgment in which researchers identified the features of language arts activities that influenced teacher judgment about the quality and usefulness of these activities. The features identified by teachers as influential were grouped under four headings: features relating to (1) students, (2) subject matter, (3) the teacher, and (4) learning environment. Of the four categories, features relating to student motivation and involvement were mentioned most frequently.

Identifying Cues for Use in Studies of Teacher Judgment

Christopher M. Clark,
Robert J. Yinger, and Susan C. Wildfong*

The most frequently used method of studying and representing judgment processes is policy capturing (Slovic & Lichtenstein, 1971; Shulman & Elstein, 1975). This approach begins with a simple model (usually linear) and attempts to reproduce the inferential responses of a particular judge. Of central interest in this paradigm is how judges weigh and combine information provided by discernible cues in the judgment task. An important first step in policy-capturing studies, then, is the identification and definition of cues to be used in the judgment task.

Since policy capturing employs linear analysis methods such as multiple regression and analysis-of-variance, the complete specification of all important cues contributing to a judgment is crucial. If some important cues are not specified in the judgment task, their effect is merely represented as a part of the predictive error or residual variance in the model. Einhorn (1974) has suggested a method for determining the functional value, i.e., the addition to predictive efficiency, of a cue or set of cues, by calculating the correlation between the residual variance of the model and the prediction criterion.

---

*Christopher M. Clark is coordinator of the Study of Teacher Planning at the Institute for Research on Teaching. Robert J. Yinger, former IRT research associate, is an assistant professor of education at the University of Cincinnati. Susan C. Wildfong is an IRT research intern.

Accurate cue specification is important in all policy-capturing studies, but it becomes a major problem in naturalistic studies. For this reason, cue specification has become a topic of particular interest to researchers interested in teacher judgment. In clinical (as contrasted with laboratory) situations, the information available to the judge is seldom clearly defined. Laboratory situations present information in a decomposed manner using individual cues at controlled and prespecified values. In the field, the judge searches for information, forms new cues from smaller ones, and uses some cues interchangeably. Also, in naturalistic studies there is always the possibility that the judge can perceive and use cues not observed by the researcher.

Unfortunately, the literature describing the methodology of policy-capturing provides little guidance for identifying, selecting, and defining cues or features of the objects to be judged. In effect, the problem is one of researcher judgment. Reflection on this problem has led us to consider alternative ways of generating cues to be used in a policy-capturing study.

Current Approaches to Cue Identification

We will discuss five approaches to identifying cues for policy-capturing studies. These strategies vary in the degree to which they are grounded in the natural judgment situation. These approaches are: (1) logical specification, (2) expert opinion, (3) prespecification and narrowing of a large number of potential cues, (4) allowing cues to emerge during a judgment task, and (5) participant observation in a naturalistic setting. Each of these approaches is discussed below.

Logical Specification

Logical specification refers to the selection of cues based on a theory pertaining to the judgment of interest or to a selection of cues from findings of previous research. In the first case, there may be theoretical work suggesting important elements to be considered in a specific judgment. Hoffman (1960) used this approach when he had individuals judge the "sociability" of 150 persons on the basis of profiles containing scores on eight selected variables from the Edwards Personal Preference Scale (Edwards, 1954). In the same study, judges were asked to rate the intelligence of 100 persons on the basis of nine cues commonly associated with this judgment (e.g., high school grades, mother's education, study habits, emotional anxiety, social status). Although the primary purpose of Hoffman's study was to examine the way in which information is used in decision making, it does illustrate one form of the logical specification method of cue generation.

The second method of logical specification is the selection of cues based on previous research. If literature exists that bears on which cues or features of objects to be judged are often used by judges, then this literature constitutes a set of nominations for cues to be employed in future studies. If hardly any such literature exists, as is the case in research on teacher judgment, then literature that deals with the objects to be judged and their functional relationships to desirable outcomes serves as the source of cues. An example of the latter approach is a study by Anderson (1977) in which she reviewed the literature on teacher effects to generate a list of teacher characteristics that correlate

with student achievement. She used these cues to vary systematically descriptions of effective and ineffective teachers. These hypothetical teachers were then judged as effective or ineffective by experienced high school teachers. This same method was used recently in studies of teacher judgments pertaining to classroom organization (Borko, 1978), classroom management (Cone, 1978), instructional strategies (Russo, 1978), and instructional content (Floden, 1978).

Logical specification is most appropriate for identifying cues in policy-capturing studies designed to test specific hypotheses or theoretical models of human judgment. In such studies it is important that the objects to be judged vary systematically along the dimensions of special theoretical interest to the investigator. This approach is less suitable for use in studies concerned with prediction and description of judgment in naturalistic situations because theoretically constructs may not accurately reflect cues actually used by judges in natural situations. An additional drawback to this approach is that, in drawing on a body of literature for cue identification, the researcher must rely on previous researchers' judgments, often not knowing the basis for their selections and definitions.

Expert Opinion

Expert opinion, the second approach to cue selection and definition, is one that has been widely used in policy-capturing studies. In this approach, experts are consulted to determine their nominations for important cues thought to be used in the judgments of interest. A recent example of this strategy is a study reported by Hammond and Adelman (1976). In this study, judgments were made about the desirability

of different types of bullets being considered for use by the Denver Police Department. To determine the important factors (cues) in judgments of the stopping effectiveness of various bullets, a panel was assembled including experts in firearms, ballistics, and medicine. Once these cues were determined, over 80 different bullets were analyzed. By combining these judgments with social value judgments, a final decision was made.

Other examples of reliance on expert opinion to identify important judgment cues include studies of physicians (Hoffman, Slovic, & Roher, 1968), pathologists (Einhorn, 1974), and stockbrokers (Slovic, 1969). Einhorn (1974) lists the conditions necessary for judgment to be considered "expert."

Expert opinion is a useful basis for cue identification in studies that attempt to model the judgment of experts or to train novices to attend to the same cues that experts report they use in coming to judgments. One difficulty with this approach is that experts in a particular area may identify cues that are desirable to use in theory, but are not actually used in practice. An example of this situation is discussed in the section below.

Prespecification and Narrowing of Potential Cues

Prespecification and narrowing of a large number of potential cues is the third approach to defining cues for policy-capturing research. There are several methods used for narrowing a large, prespecified list of cues. In Slovic's (1969) study of stockbrokers' decision making, the cue list was narrowed by means of regression analysis. Through discussions with stockbrokers, Slovic identified 11 cues he felt were potentially important in stockbrokers' decision making. He then varied these cues factorially and had two judges

make decisions on the resulting stock profiles. The two decision makers each rated 128 standardized descriptions of the stocks on a 9-point preference scale.

An interesting finding in Slovic's study was that neither decision maker appeared to be using more than half of the available attributes in the judgment task. This finding raises a question about whether or not Slovic initially listed all the attributes relevant to the decision. Wilcox (1972), commenting on the study, concluded that there is no way to judge from Slovic's results whether or not he included too few relevant cues, since the method provides no corrective signals when relevant attributes are lacking. Wilcox attributes this to Slovic's use of artificial alternatives that did not correspond to known real alternatives. The only information about the hypothetical stocks available to the decision makers was expressed in terms of the 11 attributes selected by Slovic. Therefore, it is not surprising that the decision makers' preferences were highly correlated with the given attributes. One way to test whether or not the specified cues were really being used by the judges would be to have the stockbrokers rate stocks with which they were already familiar. In this way, the cues would emerge from a real-life judgment task rather than an artificial one. Any sharp reduction in the explanatory power of the model provides a signal of possible misspecification of relevant attributes.

This discussion raises two important weaknesses of the prespecification and narrowing method. The first is that it is relatively inefficient. Slovic reported that his two judges spent about 10 hours on each task. The difficulty of the task is due to the need to cross

factorily all the relevant cues. The second weakness of this method, which is also inherent in the two previously discussed approaches, is the tendency for the researcher to project his own perceptions of the cues onto the task. This is especially true in the first step, where the researcher, through observation and discussion of cues with the judges, decides which cues appear to be relevant to the situation.

A second example of prespecification and narrowing is the semantic differential method. This method is based on factor analysis of several judges' ratings of objects on a large number of pre-specified adjective scales. The objects are then rated by summarizing attribute factors revealed by a factor analysis of ratings by all judges on all scales. By comparing the factor scores of the objects to an individual judge's rating, a measure is obtained of each individual's cue usage.

The major advantage of this method is that it is more efficient than regression analysis alone, due mainly to the reduction of the number of factors with which each judge has to deal. Because individual differences between judges' use of cues are obscured by the combination of many judges' ratings prior to the factor analysis, this approach seems more suitable for use in research in which general population characteristics and trends are of more interest than case-study descriptions of a single judge's policy.

Allowing Cues to Emerge During a Judgment Task

Allowing cues to emerge during a judgment task is the fourth commonly used approach to selecting cues. Multidimensional scaling and Kelly's Role Repertory Test (Kelly, 1955) are the two most

prevalent methods falling into this category.

Multidimensional scaling uses estimates or comparisons of inter object similarities to build up a spatial configuration of objects in which similarities correspond to inter object distances. This configuration is then analyzed, and the minimum number of dimensions in which the configuration may be embedded is determined. Wiggins (1973) describes this method in detail and provides several examples of its use.

The two primary advantages of this method are that: (1) there is no need to pre-specify the cues since they are elicited from the judge during the similarity comparison task, and (2) only weak ordinal assumptions regarding the types of comparisons of similarity used by the judges are required. The method does, however, have three major weaknesses. First, a large number of similarity comparisons is required in order to construct the stable metric of the cue dimensions. This is due to the necessity of making all possible triad comparisons using the objects selected for the task. Second, comparison of triads often requires extra work for the judge in terms of calculating combinatorial weights. Since in a triad comparison two objects must be compared against a third, the two objects must somehow be weighted in order to determine their importance. Third, it is difficult to interpret the dimensions arrived at after the task since introspective information is not elicited. In other words, cue labels are not asked for.

A method that avoids some of these difficulties is Kelly's Role Repertory Test (REP Test). This procedure generally involves four steps. First, the judge is asked to match a given list of appropriate "role" descriptions with appropriate objects from the judge's own

experience. Second, a limited number of triads of these objects is selected and the judge is asked which pair of the triad represents the best match, in what way the objects are similar, and in what important ways the third member of the triad differs. Third, the judge positions each object on each relevant attribute scale. Objects are then scored as either +1 (similar) or -1 (different) on each of the raw cue dimensions implicitly defined in the first step. Finally, these attribute data are factor analyzed to eliminate redundancies.

An advantage of using the REP Test for cue identification is that the comparison task is simplified for the judge (compared to that of multidimensional scaling) since individualized, self-selected, familiar objects are used. An additional strength of this method is that cue labels used by the judge are elicited during the task. After these labels are elicited, the similarity comparison task is cut short, further reducing task difficulty for the judge.

The method used by Wilcox (1972) in his stock market participant study was a variation of the REP Test. As with Kelly's method, a two-stage data collection procedure was used. In the first stage, called the Stock Role Repertory Exercise, a list of 20 roles that various stocks play in the subject's experience and conceptual structure was prepared (e.g., a very popular stock, the stock in which he first made a considerable gain, the stock sold too soon). Next, each judge was asked to designate a particular stock for each of these roles. Twenty triads of these stocks were then selected and presented to the decision makers. Following this, the judges were asked in what important way two members of the triad differed from the third member (as in Kelly's triad comparison task). This step was used to elicit important conceptual dimensions being used by the judge. (It is at this point in the

method that considerable experimenter judgment was called for. Wilcox reported that there was a 30% reduction of labels made at this point as he tried to eliminate redundancies.) After the relevant cues had been determined, a questionnaire was constructed asking the judge to do the following for each attribute elicited in the previous step:

1. Divide the scale into two to nine equivalent intervals.
2. Place any appropriate stocks into two separately provided categories— "scale not relevant" and "not enough information."
3. Place the remaining stocks on the attribute scale at their appropriate intervals.

The second step of Wilcox's method involved a factor analysis (using principal component analysis) of the questionnaire data in order to condense the raw attribute factor structures.

Wilcox's method has several strengths, notably: The factor analysis using the data for a single decision maker eliminates most of the previously-mentioned difficulties of the semantic differential. Second, by allowing the judge to divide the scale into two to nine equivalent intervals much more ordinal or metric information is provided on each attribute scale. This is in contrast to Kelly's method and the multidimensional scaling method that make the assumption of equal scale intervals for all attributes. Allowing cues to emerge during a judgment task is a useful strategy in research that is primarily descriptive as contrasted with research aimed at testing specific hypotheses or theoretical models of the judgment process.

Participant-Observation

Participant-observation in the naturalistic judgment setting is the fifth method for generating significant cues for a policy-capturing

study. This approach involves careful observation of naturally occurring instances of the judgments of interest. The observer immerses himself or herself in the judgment situation, attempts to understand the frame of reference of the judge, and, indeed, becomes a judge. Identification of important cues or features of the objects to be judged then becomes not only a matter of observation and interrogation but also a matter of introspection. A method similar to this was used by Yinger (1977) to identify important cues used by an elementary school teacher in her instructional planning and judgments and decisions. Although the identified cues were used to develop a more general decision model than is usually the case in policy-capturing studies, the information collected could have easily been used as a basis for a policy-capturing analysis of a specific instructional judgment.

The major strength of this method is that there is a high probability that the judgment cues identified are ones being actually used by the judge in a natural context. The long time involved in participant-observation, however, usually precludes using this approach, unless the information is gathered as part of a larger study. The participant-observation approach is most appropriate for exploratory and descriptive research.

Summary

The goal of cue identification, selection, and definition prior to a policy-capturing analysis of judgment is to identify accurately the important cues used or usable by judges in coming to their judgments. This task is essentially one of researcher judgment as the researcher tries to ground the process on reliable and accurate information. The five approaches to cue identification vary in

practical complexity and in the extent to which they are grounded in theory or in actual naturalistic judgments. Some approaches are more appropriate for use in studies concerned with the testing of specific hypotheses about judgment processes while others seem more useful for description of cue utilization and cue salience for groups or individual judges in naturalistic situations. Synthesis and analysis are involved at different points in each method of cue identification, sometimes primarily by the researcher, in other cases primarily by the judges themselves.

An additional consideration for the researcher on human judgment is whether it is more useful to employ the same or different judges in the cue-identification and policy-capturing phases of a study. If the primary purpose of the research is ideographic, i.e., the precise description of the judgment policies employed by a small number of judges, then it probably makes good sense to use the same judges in both phases of the research. If the research is more concerned with generalizing from the observer judgment policies of a few judges to the processes employed by a larger population of similar judges, it might be more advisable to have different individuals participate in each phase.

The remainder of this paper is devoted to presenting and illustrating an approach to cue identification and definition that is a variation on the method of allowing cues to emerge during a judgment task. This variation is less quantitative than either multidimensional scaling or Kelly's method, but it allows for a specification of judgment cues used by a group of individuals in language coinciding with their own cue labels. This approach will

be illustrated in a study of teacher judgments of language arts materials.

A Study of Teacher Judgment

The purpose of our study was to identify the features of language arts activities that influence teacher judgment about the quality and potential usefulness of these activities.

Method

Subjects. Thirteen elementary school teachers, who were students in a graduate education class, were volunteers in the study. Two were male and 11 female. The average number of years of teaching experience was 4.5, with a range from three to eight years of experience. Eight teachers taught in self-contained classrooms, while five taught in team-teaching situations. Two of the teachers taught in urban settings, two in suburban communities, and nine in small towns in rural or semi-rural areas.

Procedure. The teachers were each given a booklet containing 26 descriptions of language arts activities on writing selected from a commercially available instructional catalog (Forte, Frank, & MacKenzie, 1973). The teachers were asked to rate each activity as either high, medium, or low in attractiveness for use with their most recent group of students. The following instructions were given to each teacher:

> You are about to begin a process of judgment of language arts activities. The first stage consists of rating the attractiveness of each activity on a three-point scale (high, medium, or low).

A rating of high attractiveness means that you like the activity very much, you would use the activity in classroom teaching, and you predict that the activity would be an effective learning experience for students.

A rating of low attractiveness means that you do not like the activity at all, that you would not consider using the activity in classroom teaching, and that you predict that the activity would not be an effective learning experience for students.

A rating of medium attractiveness means that you judge the activity to be neither high nor low.

Read the description of the first activity. Then reflect for a moment on your judgment of the activity. Record your judgment of this activity on the activity judgment record provided. Mark only one judgment for each activity (high, medium, or low). Repeat this process with each activity in turn.

After rating all the activities, the teachers were asked to re-examine each activity rated high. They were then asked to list the features of each "high" activity that contributed to their high rating of the activity's attractiveness. This procedure was then repeated for the activities rated low in attractiveness. The entire judgment task was completed by the teachers in one two-hour session.

To identify the important features that influenced the judgments of this group of teachers the following procedure was used:

1. Each activity feature identified by the teachers was typed on a 3 x 5 card; 440 of these statements were produced.

2. Two of the researchers independently categorized each of the 440 teacher statements.

3. The two researchers compared their categorization of the teacher statements, clarified their own policies for categorization, and negotiated a final set of 13 categories into which 407 teacher statements could be sorted. The remaining 33 statements were unclassified because they did not appear to relate to the judgments of activity attractiveness.

Results

Results of this study are presented below in three sections. The first section deals with the range and variety of cues identified by the teachers as being influential in their judgment. The second section addresses individual differences in the number of features cited by each teacher. The third section deals with the configuration of features that characterizes language arts activities rated highest and lowest in attractiveness.

Features identified. The features of language arts activities identified by the teachers as influential in their judgment are listed in Table 1. These features have been grouped under four headings: features of activities that relate to students, the subject matter, the teacher, and the learning environment.

Table 1
Features of Language Arts Activities Identified as Influences on Teacher Judgment

Feature Category	Frequency of Report
STUDENTS	
Motivation and Involvement	104
Cognitive Outcomes	67
Affective Outcomes	44
SUBJECT MATTER AND MATERIALS	
Difficulty	68
Fit between Purpose and Process	35
Meaningfulness	16
Tangible Product	15
Clarity	11
Integration	6
TEACHER	
Fit with Style	11
Demand on Teacher	8
Benefit for Teacher	3
ENVIRONMENT	
Fit with Behavior Setting	19

In this judgment exercise, the teachers most frequently mentioned activity features that were related to student behavior. Student motivation and involvement was mentioned most frequently as a basis for accepting or rejecting a language arts activity, followed by features of activities thought to influence cognitive student outcomes and affective student outcomes.

In the subject matter category, the feature most frequently singled out was the estimated difficulty of the activity for that teacher's class. The frequency with which this feature was mentioned was probably inflated due to the fact that the language arts activities were intended for students in the upper elementary grades, while four of the teachers taught primary-grade students.

Half of the activity features relating to the teacher concerned the match between the activity description and the teacher's prediction of how comfortable he or she would be in executing this activity. The remaining features in this category had to do with how difficult the activity would be to set up, manage, and evaluate, and whether the activity benefited the teacher in addition to its stated purpose.

The final category of features identified had to do with how well the activity seemed to fit into the particular classroom organization that the teacher was working with. For example, some activities were judged to be particularly appropriate for use in learning centers.

Individual differences. Teachers differed in both the number of features mentioned as influencing their judgment and in their distribution of features across the 13 categories listed in Table 1. Table 2 displays these data for indivudual teachers. As to number of different features reported, teachers ranged from using only

three features (Teacher 11) to as many as 11 features (Teacher 12). The average number of activity features mentioned as influencing teacher judgment was 7.8.

As can be seen in Table 2, the frequency distributions of features mentioned vary widely across teachers. All teachers mentioned student motivation and involvement as an influence on their judgment, but this is the only activity feature used unanimously. Some teachers seemed primarily concerned with student cognitive outcomes (Teachers 4, 7, 10, and 13), while Teacher 6 attributed her judgments largely to affective student outcomes. Although all but one of the teachers mentioned difficulty of the material at least once, Teacher 11 stands out as particularly sensitive to this feature. This can be explained by the fact that Teacher 11 taught first grade, while the language arts activities used in this study were intended for upper elementary students.

The fit or match between the stated purpose of an activity and the process whereby it was to be carried out was another feature mentioned by all but one of the teachers, although it was mentioned less frequently than difficulty or any of the three student-related features. The remaining subject matter and material, teacher, and environment-related features were used to explain teacher judgments much less frequently. However, an examination of Table 2 does indicate that for certain teachers some of these other features did figure substantially in their comments.

Table 2
Frequency of Reported Use of Activity Features for Each Teacher

	FEATURE													
	Students			Subject Matter and Materials						Teacher			Environment	
Teacher	Motivation & Involvement	Cognitive Outcomes	Affective Outcomes	Difficulty	Fit Between Purpose and Process	Meaningfulness	Tangible Product	Clarity	Integration	Fit With Style	Demand on Teacher	Benefit for Teacher	Fit with Behavior Setting	Number of Categories Reported
1	13	2	1	3	4	-	-	4	-	-	-	1	6	8
2	8	5	4	8	3	1	3	-	-	2	5	-	-	9
3	7	6	1	3	3	1	-	1	-	1	-	-	-	8
4	6	11	4	4	5	6	-	1	1	-	-	-	2	9
5	12	2	2	2	2	-	1	-	-	-	-	-	1	7
6	9	-	10	8	1	1	1	1	1	1	-	-	1	10
7	7	10	1	8	3	-	-	-	1	1	-	-	-	7
8	5	2	1	6	1	-	-	1	1	2	1	-	-	9
9	5	5	-	4	1	-	-	1	-	-	-	-	-	5
10	12	14	9	-	4	2	6	-	-	-	-	-	4	7
11	4	-	-	18	-	-	1	-	-	-	-	-	-	3
12	13	1	6	3	2	-	2	2	1	2	2	-	5	11
13	4	9	5	1	6	5	1	-	1	2	-	2	-	10
TOTAL	104	64	44	68	35	16	15	11	6	11	8	3	19	

Feature Configuration

The third method for analyzing these data was to examine the configuration of features characterizing activities rated highest and lowest overall. An analysis of the teacher comments concerning the four activities rated highest by all teachers and the five activities rated lowest indicated that the feature mentioned most frequently as contributing to positive teacher judgments was student motivation and involvememt (30 comments or 62.5%). The second most frequently cited influence on positive teacher ratings was difficulty level of the activity (10 comments or 20.8%). The third most frequently mentioned feature was the fit between the purpose of the activity and the process of teaching involved in the activity (7 comments or 14.5%). In sum, activities rated highest by the teachers were those high on student motivation and involvement, low in difficulty, and those that were perceived as good ways to teach the content.

In the case of the five activities rated lowest, the most powerful influence identified by the teachers was the difficulty of the activity for students (13 comments or 48.1%). A second feature that influenced teachers to reject these activities was the amount of demand that the activity placed on the teacher (7 comments or 25.9%). In other words, activities were rated as unattractive if they were seen as too difficult for the students or too demanding for the teacher.

Discussion

The results reported above describe the process and products of a method for identifying cues for use in studies of teacher

judgment. We believe that, with minor modifications, this approach to cue identification can be useful in the preparatory stages of descriptive studies of human judgment in areas other than teaching. The method in this study appears to possess the advantages of simplicity and realism.

Simplicity. The method used in this study involves a relatively simple task for the participating judges; the total time required to make judgments of attractiveness of 26 language arts activities and to record features of those activities that influence teacher judgment was 1.5 hours. The researchers spent approximately 10 person-hours categorizing the activity features so identified. These data compare very favorably with the time requirements required by other approaches to cue identification (e.g., Slovic, 1969; Wilcox, 1972). In some of these studies, individual judges spent as long as 10 hours specifying cues or making diadic and triadic comparisons of objects to be judged. Estimates of the time required to reduce and analyze such data are not available in these original reports, but one can imagine that the time spent was as lengthy as the data were complex.

Realism. Perhaps the strongest feature of this approach to identifying cues is that the cue-identification task very closely approximates the real-world judgment situation of interest. Teachers do examine instructional materials, decide which ones are attractive or potentially useful, and can articulate reasons for these judgments. This is in contrast to methods in which the judge enters a hypothetical situation with which he or she has had no experience, making and justifying judgments that are not typical of the kinds made in non-experimental situations.

A second aspect of realism in this approach is that the language of the judges is preserved as much as possible in the generation of the final cue categories. Since the researchers form categories by sorting and combining actual statements made by judges, the resulting cues are likely to be strongly influenced by the terms in the judges' lexicon.

Limitations

The two primary limitations of the method described here are that researcher judgment is still a potentially weak link in the cue identification process and that there are unknown limits on a judge's ability to report veridically those features of an object to be judged that actually influence judgment. In the present case, researcher judgment was involved in selecting the initial array of language arts activities to be judged and, more critically, in grouping and classifying individual teacher reports of cues used. Although an attempt was made to capture the flavor and terminology used by the 13 teachers as we generated cue labels, the final set of cues reflects the mental category systems of 13 teachers and two researchers in unknown proportions. It would have been interesting to have had a teacher participate in the cue grouping, categorization, and labeling exercise. Unfortunately, this was not done.

The second limitation of this method is that the validity of the cues reported by teachers is unknown. Nisbett and Wilson (1977) argue that human judges sometimes attribute their decisions to factors that they could not have possibly been aware of at the time that the judgments were rendered. We take the less radical position that at least some reasonably large fraction of a judge's reported basis

for judgment can be validly recalled. Further, the present approach requires the judge to feature-analyze the objects to be judged immediately after the judgment has taken place. This differs from the expert-opinion approach in that the judgment exercise is grounded in an actual judgment task and that multiple practicing judges' judgments are brought to bear in the generation of cues.

Conclusion

In summary, a method of identifying judgment-relevant cues has been proposed that is quite different from those used in current practice. While the method described and illustrated above does not solve all of the technical, psychological, and philosophical problems involved in cue generation, it does offer a technique that is relatively simple to use with a group of judges, has relatively high ecological validity (Snow, 1974), and invites serious attention to the importance of the cue generation phase of the study of human judgment. More attention must be paid to the problem of cue identification if realistic and useful research on human judgment in complex natural settings is to be done.

References

Anderson, B.L. Differences in teachers' judgment policies for varying numbers of verbal and numerical cues. Organizational Behavior and Human Performance, 1977, 19, 68-88.

Borko, H. An examination of some factors contributing to teachers' preinstructional classroom organization and management decisions. Paper presented at the annual meeting of the American Educational Research Association, Toronto, Ontario, March 1978.

Cone, R. Teachers' decisions in managing student behavior: A laboratory simulation of interactive decision-making by teachers. Paper presented at the annual meeting of the American Educational Research Association, Toronto, Ontario, March 1978.

Edwards, A.L. Edwards personal preference schedule. New York: Psychological Corp., 1954.

Einhorn, H.J. Cue definition and residual judgment. Organizational Behavior and Human Performance, 1974, 4, 30-49.

Floden, R.E. A policy-capturing study of teacher decisions about content. Paper presented at the annual meeting of the American Educational Research Association, Toronto, Ontario, March 1978.

Forte, I., Frank, I., & MacKenzie, J. Kids stuff: Reading and language experiences, intermediate - jr. high. Nashville, Tenn.: Incentive Publications, 1973.

Hammond, K.R., & Adelman, L. Science, values, and human judgment. Science, 1976, 194, 389-396.

Hoffman, P.J. The paramorphic representation of clinical judgment. Psychological Bulletin, 1960, 57, 116-130.

Hoffman, P.J., Slovic, P., & Roher, L.G. An ANOVA model for the assessment of configural cue utilization in clinical judgment. Psychological Bulletin, 1968, 69, 338-349.

Kelly, G.A. The psychology of personal constructs: A theory of personality (Vol. 1). New York: W.W. Norton and Company, Inc., 1955.

Nisbett, R.E., & Wilson, T.D. Telling more than we can know: Verbal reports on mental processes. Psychological Review, 1977, 84, 231-259.

Russo, N.A. Capturing teachers' decision policies: An investigation of strategies for teaching reading and mathematics. Paper presented at the annual meeting of the American Educational Research Association, Toronto, Ontario, March 1978.

Shulman, L. S., & Elstein, A.S. Studies of problem solving, judgment, and decision making. In F.N. Kerlinger (Ed.), Review of Research in Education (Vol. 3). Itaska, Illinois: F.E. Peacock and Co., 1975.

Slovic, P. Analyzing the expert judge. Journal of Applied Psychology. 1969, 53, 255-263.

Slovic, P., & Lichtenstein, S. Comparison of Bayesian and regression approaches to the study of information processing in judgment. Organizational Behavior and Human Performance, 1971, 6, 647-744.

Snow, R.E. Representative and quasi-representative designs for research on teaching. Review of Educational Research, 1974, 44, 265-291.

Wiggins, N. Individual differences in human judgments: A multivariate approach. In L. Rappoport & D.A. Summers (Eds.), Human judgment and social interaction. New York: Holt, Rinehart & Winston, Inc., 1973.

Wilcox, J.W. A method for measuring decision assumptions. Cambridge, Mass.: MIT Press, 1972.

Yinger, R.J. A study of teacher planning: Description and theory development using ethnographic and information processing methods. Unpublished doctoral dissertation, Michigan State University, 1977.